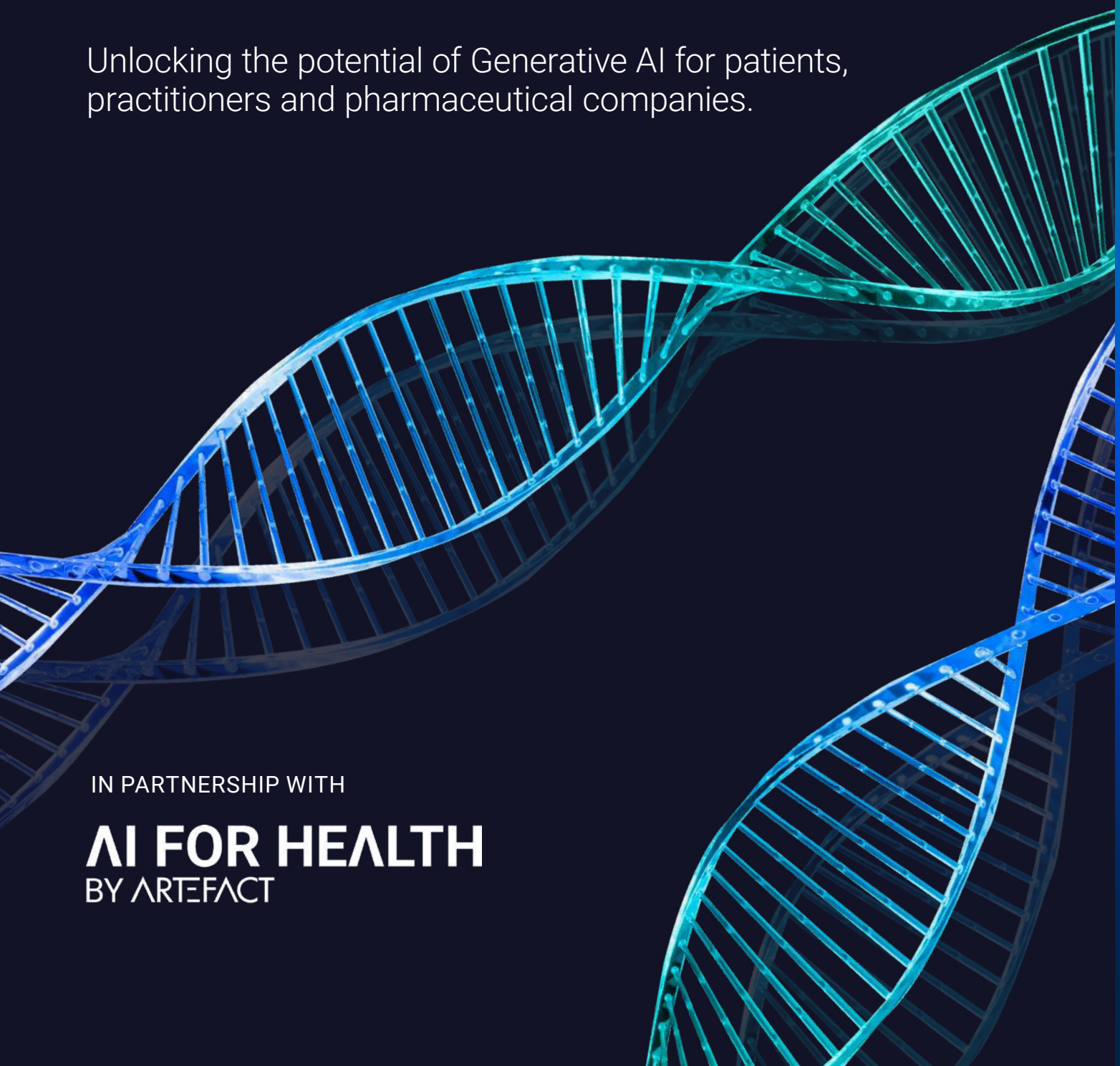
ARTEFACT

GENERATIVE AI REPORT FOR HEALTHCARE

Unlocking the potential of Generative AI for patients, practitioners and pharmaceutical companies.



IN PARTNERSHIP WITH

AI FOR HEALTH
BY ARTEFACT

ARTEFACT

DATA IS ABOUT PEOPLE

WE UNLEASH THE FULL VALUE OF YOUR PEOPLE AND BUSINESSES THROUGH DATA AND AI SOLUTIONS.

18
COUNTRIES

+1300
EMPLOYEES

+1000
CLIENTS

Artefact is a global leader in consulting services, specialized in data transformation and data & digital marketing, from strategy to the deployment of AI solutions. We offer a unique combination of innovation (Art) and data science (Fact).

DATA ACCELERATION PROGRAMS | AI SOLUTIONS | DATA MARKETING

GENERATIVE AI REPORT FOR HEALTHCARE

Unlocking the potential of Generative AI for patients, practitioners and pharmaceutical companies.

1 **The rise of Generative AI in healthcare: Navigating between promise and control** 04

2 **Concrete Generative AI healthcare: Applications and benefits** 07

3 **The healthcare ecosystem: Gearing up to unlock the potential of Generative AI** 18

4 **Generative AI in healthcare: Limitations, challenges and opportunities** 33

5 **Trust and control: A critical role in realizing the potential of Generative AI in healthcare** 40

1 FOREWORD

The rise of Generative AI in Healthcare: Navigating between promise and control

Generative AI has experienced massive adoption and has gained significant momentum in 2023. While its prospects for healthcare are promising, the biggest challenge will be our ability to control outcomes.

We are witnessing unprecedented hypergrowth in medical knowledge: while it doubled every 50 years in the 1950s, it accelerated to 3.5 years in 2010 and then to every 73 days in 2020. Knowledge is expanding faster than our ability to consume it, both in patient care and research. At the same time, the patient population is growing and complex pathologies are becoming more common, leading healthcare professionals toward hyperspecialization.

These phenomena put significant pressure on healthcare professionals, who are overwhelmed by the content while having less time to assimilate it. Therefore, it is crucial to leverage technology to support them.

Generative AI represents a profound paradigm shift in the way we work today, as models switch from being task-focused to outcome-driven. Thanks to foundational models, machines are now making an effort to understand humans and transform the way they consume data.

Here is a short list of interesting prompts that healthcare professionals might consider using to enhance their daily activities:

Generative AI holds the transformative potential to liberate humans from repetitive tasks, allowing them to focus their efforts on higher-value activities and freeing up time to address more complex needs.

- **It can automatically summarize data regardless of volume.**
- **It facilitates output customization, allowing to refine results, modify their form, language...**
- **It supports the creative process by quickly generating ideas and different alternatives of the same result.**

The impact on the industry is estimated at $1 trillion (McKinsey: Tackling Healthcare's Biggest Burdens with Generative AI). We foresee impacts across the entire healthcare value chain, which are developed in the first section of this white paper.

However, these opportunities come with significant challenges. If we set aside the traditional issues we face with AI (ethical, regulatory...), two challenges remain.

The first is the potential for Generative AI models to hallucinate. Most large language models are based on a specific neural network technology called a transformer. This model is trained to identify patterns, to make connections between concepts, to predict words one after another. This partly explains hallucinations: nonsensical sentences of completely fabricated information. This is a major concern for some applications that directly affect the health of patients (e.g., diagnostic support, patient information...). Therefore, a key success factor for Generative AI applications in healthcare will be the validation and control processes.

The second is complementary to the first: these hallucinations appear so plausible that it becomes very challenging to distinguish between what's true and what's false. For example, Chat GPT results often appear to be perfect in form, well written and presented, with arguments that seem almost irrefutable - but are not necessarily true. Results like these naturally inspire trust for end users who have not been forewarned. Therefore, the second key success factor for the application of Generative AI will be the training of healthcare professionals and patients. Every end user should know what to expect from these applications and what their limitations are.

Earlier this year, Artefact released a Generative AI survey describing the technology, the rewards and the risks. This comprehensive survey covers the applications, technical challenges, human impacts and ethical stakes across industries. Given the specificities of the healthcare industry, we decided to write this white paper in order to answer the following questions:

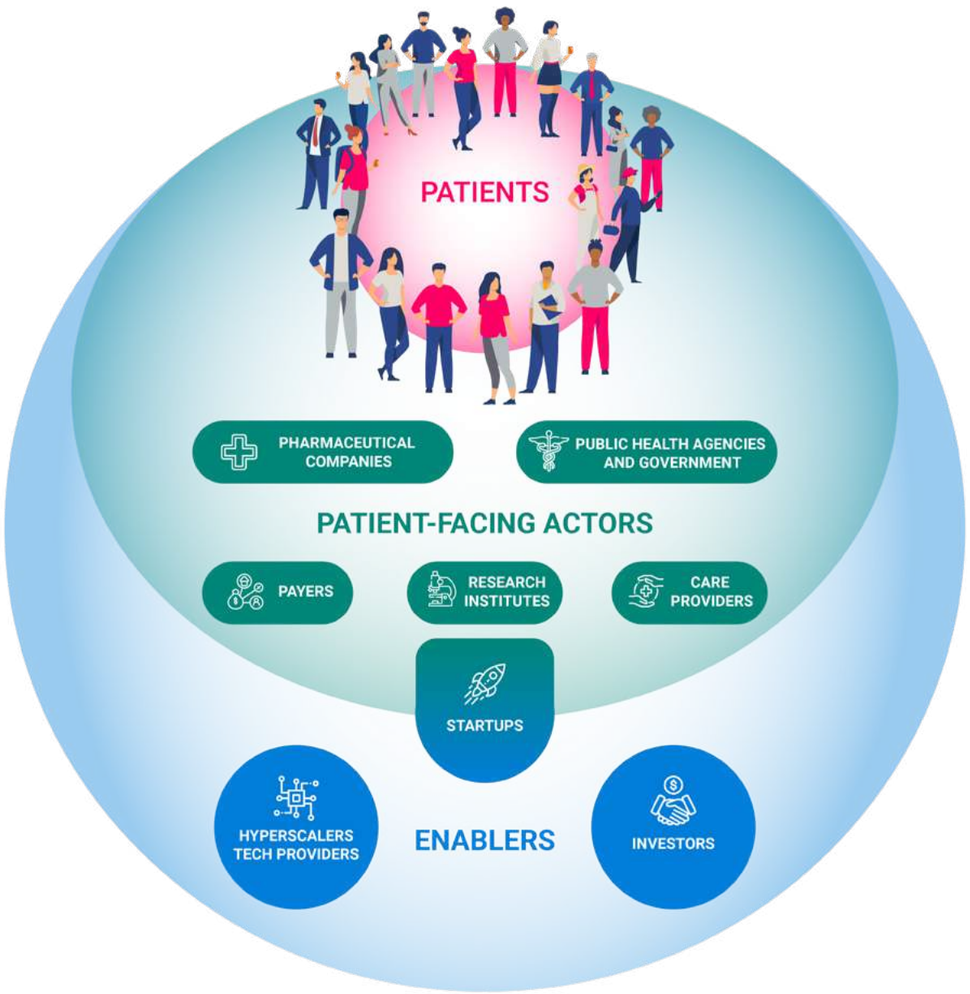
- How can Generative AI transform the healthcare industry, and what will be its first applications?
- Who are the main actors contributing to the spread of Generative AI, and what are their current areas of focus?
- What are the key challenges in implementing Generative AI applications and what current solutions are being explored?

Paul de Balincourt
Director - Healthcare Data & AI Transformation
ARTEFACT

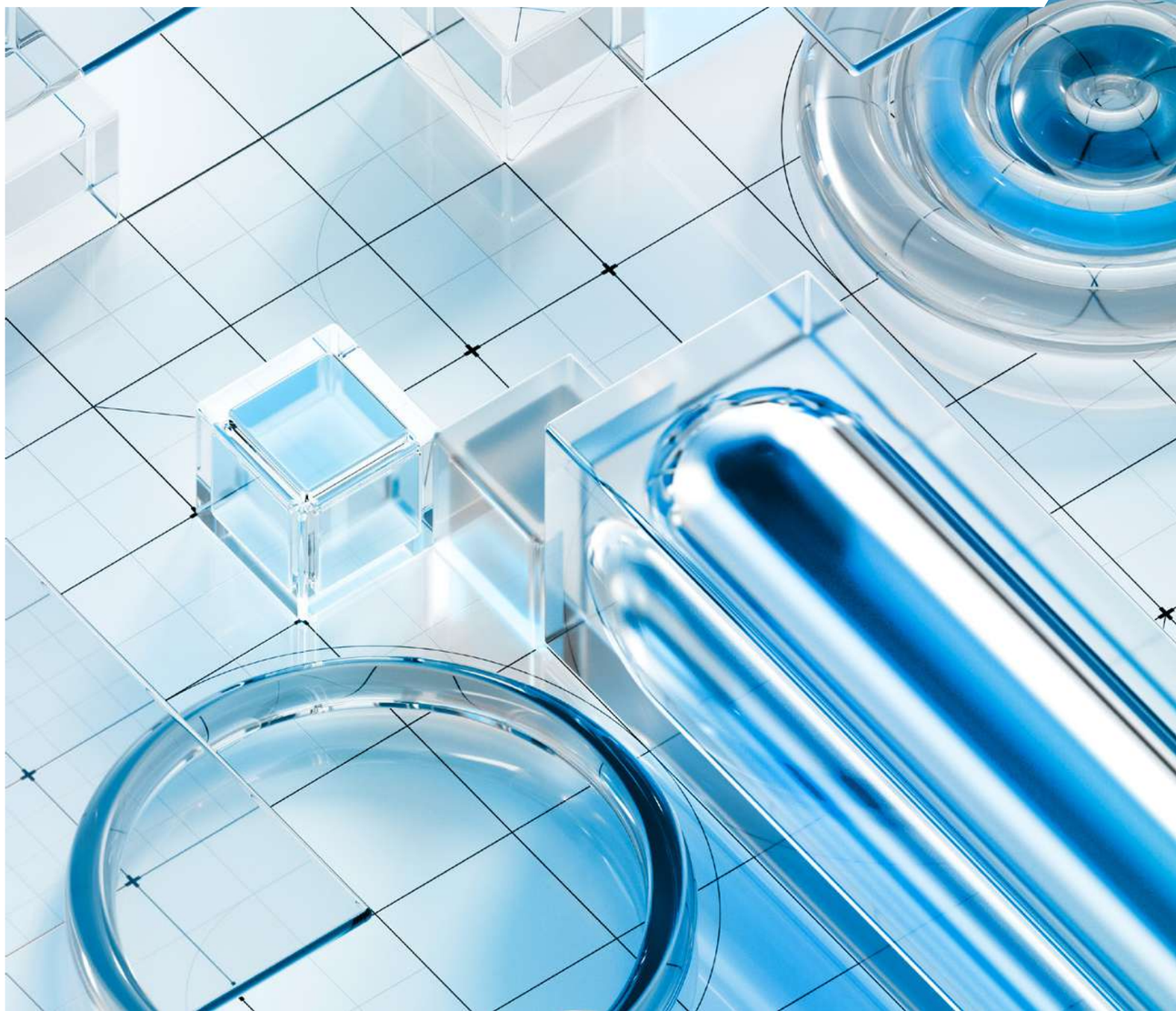
Our methodology

The aim of this work is not to be exhaustive, but rather to highlight disruptive Generative AI use cases and players within the healthcare ecosystem that have captured our interest.

To write this white paper, we took an ecosystem-focused approach and categorized the Generative AI healthcare landscape into two main groups: patient-facing actors and enablers. We chose to place the patient at the center of our analysis because we believe that the interests and safety of patients should always come first when considering the launch of a Generative AI initiative in the healthcare and pharmaceutical industries. For each type of actor, we look at current and future use cases for Generative AI, how they are gearing up to develop these applications, and the limitations, challenges, and opportunities they may face.



2 | Concrete Generative AI healthcare: Applications and benefits

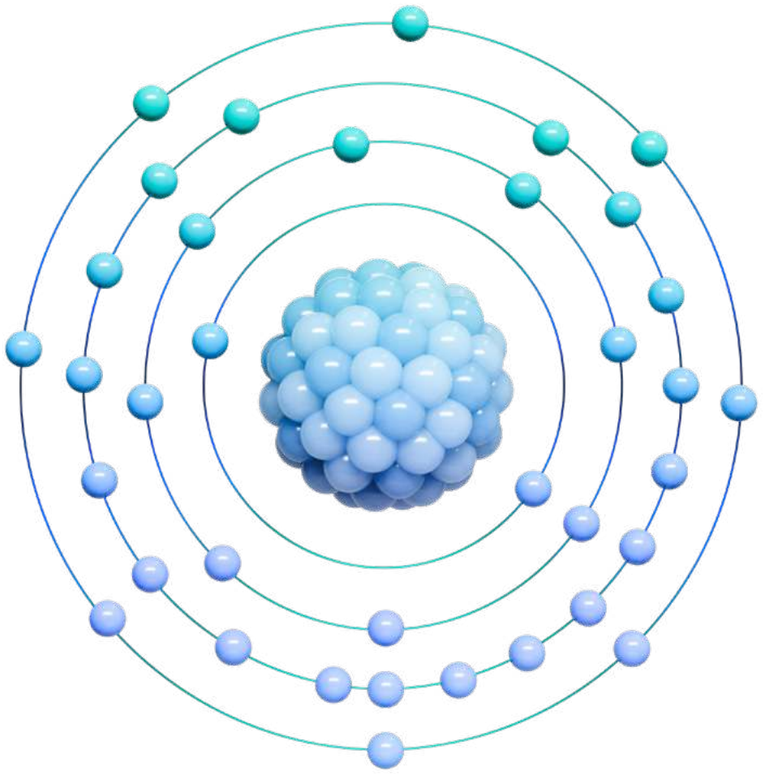


Generative AI is already impacting the entire healthcare ecosystem

Applied to the healthcare industry, Generative AI has the potential to deliver significant short-term improvements in operations management, especially through the development of virtual assistants and human augmentation for researchers and caregivers. In the medium to long term, it should drive real innovation acceleration in high value-added areas such as drug discovery, precision medicine, and care decision making.

The rapid and widespread adoption of Generative AI solutions in 2023 has made it challenging to synthesize all potential uses. Nonetheless, several prominent categories of use cases seem to emerge:

- **Data augmentation: Generate new data to enrich/ expand the datasets used to train AI-based medical devices or validate clinical trials.**
- **Insight generation: Summarize, extract key information from data to support diagnostics and drug discovery.**
- **Biomolecule generation: Design new drug candidates from scratch by generating and optimizing de novo biomolecular structures with the desired properties for drug development.**
- **Content personalization: Analyze interactions and create personalized content/experiences to enhance communication with patients and HCPs.**
- **Productivity and automation: Enhance day-to-day operational tasks such as assisting with code development or completing administrative tasks (e.g., prescription forms, office visit reports, social security forms).**



For some advanced applications of Generative AI, it is critical to keep scientific expertise at the core of the process. For example, we know that AI can be a great ally in the drug discovery process for target identification and compound design. However, in the preclinical development phase, pharmacology and toxicology studies are essential to validate the identified leads and optimize them for efficacy, safety and better drug-like properties. Biological and chemical knowledge remain at the core of development.

It is difficult to predict where and when these use cases will impact the industry. Without making predictions, we can estimate that the following factors will determine the time-to-market of these Generative AI-enabled solutions:

- **Technological maturity: The current level of readiness of technology assets (e.g., foundational models available and performant for a dedicated application).**
- **Adoption effort: Depending on the issue we are trying to address, the challenges of reliability, validation, and user acceptance vary (e.g., applications delivered directly to patients face significant adoption challenges).**
- **Data readiness: The ability to capture and leverage enough qualitative data to feed the model.**

> ***It is hard to believe that Generative AI will be able to create successful new therapeutic programs without a deep understanding of biology.***
>
> **Thomas Clozel -** CEO and Co-founder **- OWKIN**

In the example below, we propose a holistic view of the main Generative AI use cases that can be expected in the healthcare and pharmaceutical domains:

PHARMACEUTICAL COMPANIES

1 - Synthetic patient data generation for clinical trials: Use Generative AI models to reduce the number of patients to recruit (e.g., for rare diseases or for risk populations).

2 - De novo biomolecule generation: Craft new protein structures thanks to diffusion models to accelerate drug development and improve gene therapy.

3 - HCP engagement assistant for sales reps: Generate optimal responses for sales reps based on phone calls and emails to better engage HCPs.

4 - Marketing content generation: Use Generative AI applications to create different adaptations (e.g., messages, formats) of a specific marketing content.

5 - Regulatory conversational assistant: Build a tool to enable regulatory teams to interactively access regulatory resources by country and receive reminders and checklists at each step of the drug approval process.

CARE PROVIDERS

6 - Biomarker analysis support: Automate biomarker data processing (text, images) with Generative AI to enable remote patient monitoring and thereby reduce the number of invasive examinations (e.g., biopsies).

7 - Image enhancement and analysis for diagnosis and treatment planning: Enhance medical imaging by generating synthetic images, improving reconstruction and segmentation and facilitating disease diagnosis and treatment planning.

8 - Diagnosis and care decision-making support: Generate evidence-based, explainable next steps in a diagnosis or car decision-making context with relevant sources available for review.

9 - HCP administrative assistant: Manage the summarizing of new medical content (e.g., studies) and writing for administrative tasks (e.g., taking minutes of patient appointments, writing prescriptions, or updating patient records).

10 - Medical coding: Provide real-time coding suggestions and recommendations, ensuring accuracy and compliance by offering insights into appropriate codes, modifiers and documentation based on industry standards.

RESEARCH INSTITUTES

11 - Writing and formatting scientific publications: Accelerate literature review, identify research questions, provide an overview of the current state of the field, and automate formatting and language review.

12 - Patient pre-screening: Scan a multitude of health and medical records to streamline the recruitment funnel and better identify appropriate candidates for clinical trials.

PUBLIC HEALTH AGENCIES AND GOVERNMENT

13 - Public health and resource monitoring: Detect early signs of outbreaks, monitor and predict the spread of pathogens, and identify sources of infection by analyzing vast amounts of population data.

14 - Mental health support: Improve mood monitoring, detect changes in human behavior through text, images or voice, and characterize early symptoms of depression or anxiety.

PAYERS

15 - Preventive and informational agent for patients: build informational conversational agents to provide information on medical conditions, surgeries, medical procedures, medications, preventive care and share patient education material.

16 - Claims processing streamlining: Automate some routine tasks, such as data entry and analysis. Summarize and prioritize claims.

STARTUPS

Startups can address all the mentioned use cases.

Transforming healthcare: Zoom in on several high-impact use cases of Generative AI

In the following section, we present a selection of flagship use cases that are already available or in development and are likely to have a near- to mid-term impact on the pharmaceutical and healthcare industries.

NOTE

Please be advised that all content created with Generative AI is subject to a mandatory human-in-the-loop approach and rigorous assessment for potential risks and compliance with applicable standards and regulations.

USE CASE #1

Synthetic patient data generation to accelerate clinical trials

> *Generative AI can reduce the time needed for the third phases of clinical trials, thanks to 'augmented cohorts' (i.e., virtual patients generated by AI) even though physician validation is required at every stage of the process.*
>
> **Stéphanie Allassonnière**
> Professor and Vice-President
> Valorisation and industrial partnerships
> **UNIVERSITÉ PARIS CITÉ**

TARGET

Clinical Trial Investigator: As a clinical trial investigator, I oversee and conduct clinical trials to evaluate new medical treatments and interventions, ensuring they meet ethical and regulatory standards.

CONTEXT

One major hurdle in clinical trials[1] is patient recruitment. As the phase III of clinical trials usually requires a large volume of patients (thousands of patients are sometimes needed), it can take up to four years to gather a sufficiently large group[2] to launch this phase. In addition, even once recruitment is completed, it does not ensure the retention of patients[3] afterwards.

CHALLENGE

- Accelerate patient recruitment to reduce the time to phase III launch.
- Facilitate the recruitment of multicentric cohort recruitment for complex populations such as rare diseases and high-risk patients.
- Reduce potential retention difficulties.

SOLUTION

- **Centralize all past and current patient journey data[4] from your current cohort.**
- **Implement a generative adversarial network model[5] on your cohort data set to virtually augment it with synthetic patient data, including information on clinical features, genomics, treatment and outcomes.**
- **Leverage a validation framework to assess fidelity and privacy preservability of the generated data.**

RESULTS

- Solve real patient data imbalance and incompleteness issues.
- Preserve patient privacy as the generated data are neither a copy nor a representation of the real data.
- Accelerate the clinical trials process and reduce the cost of patient recruitment.

USE CASE #2

Personalized care recommendation support

> *Generative AI has the advantage of embracing the complexity of a situation; but in the end, it's up to the human to be able to make the decision ; this prevents the patient from feeling perceived only as a data set.*
>
> **Eric Vibert**
> Professor and MD, PhD, IKO, Liver Surgeon in Paul Brousse Hospital AP-HP, Paris Saclay University
> **CHAIRMAN OF BOPA**

TARGET

Medical practitioner: As a medical practitioner, I diagnose and treat medical conditions, conduct assessments, prescribe medications, and teach patients to improve their health and well-being while following ethical and medical guidelines.

CONTEXT

Medical practitioners are guided in their care decision-making by standards of care recognized by the global healthcare community and by recommendations[6] issued by global or national healthcare regulatory bodies, such as the Haute Autorité de Santé (HAS)[7] in France. Over the years, regulatory recommendations have multiplied and become more complex. For patients with multiple pathologies, it can be time consuming for doctors to navigate these numerous standards and recommendations and apply them to patient care routines.

CHALLENGE

- Support medical practitioners in care decision-making by generating evidence-based and personalized next step proposals.
- Bring care practices closer to standards of care and regulatory recommendations.

SOLUTION

- **Continuously collect all past and current patient journey data.**
- **Train an LLM on a curated data set of global and national healthcare regulatory recommendations and key scientific publications describing current and emerging standards of care.**
- **Contextualize the LLM with patient historical data and ad hoc practitioner insights.**
- **Generate evidence-based, personalized, explainable next steps in a diagnosis or care decision-making context with relevant sources available for review.**

RESULTS

- Streamline and improve global quality of care by helping physicians more systematically implement regulatory recommendations and standards of care.
- Enable medical practitioners to spend more time with patients.

USE CASE #3

Healthcare professional (HCP) administrative assistant

TARGET

HCP: As an HCP, I am a member of the medical, dental, pharmacy or nursing team. I interact with patients and I may prescribe, purchase, supply, recommend or administer medical treatments and products.

CONTEXT

As a healthcare practitioner, staying up to date with medical research is a challenge. The exponential growth of medical knowledge makes it difficult for healthcare professionals to keep up with every new piece of information. This struggle can have a direct impact on the quality of care provided to patients.

CHALLENGE

- Find useful and relevant content among numerous sources (studies, clinical guidelines, research papers…).
- Be able to manipulate and format content in order to better memorize, use and share it with other HCPs.

SOLUTION

- **Centralize qualitative medical content to make it searchable by a Generative AI model.**
- **Train the proper LLM (e.g., Med-PalM by Google) to support identified prompts (i.e., summarization, source identification, medical questions…).**
- **Use a testing and validation framework, to assess summarization quality, question answering accuracy, recommendation relevance, etc.**

> *In 2020, over 100,000 articles were published on a single pathology: COVID. Generative AI has the potential to relieve healthcare professionals who lack the time to keep up with the ever-expanding volume of scientific literature by providing them with generated summaries of publications.*
>
> **Grégoire Pigné**
> CEO - Oncologist and Radiation Therapist
> **PULSELIFE**

RESULTS

- Increased expertise, enabling healthcare professionals to effectively manage a wider range of medical conditions.
- Improved efficiency, allowing healthcare professionals to spend more time focusing on patient care.
- Improved evidence-based decision-making.

ARTEFACT

USE CASE #4

Medical coding assistant for hospitals and clinics

TARGET

Health information manager: As a health information manager, I collect, analyze, code, maintain, and secure a healthcare facility's medical information for billing, research, quality improvement, and credentialing.

CONTEXT

Medical procedures require a code[8] for transmission to insurance or social security systems. This code ensures accurate identification of the medical act for reimbursement purposes. Currently, coding is manually entered into systems, resulting in errors and omissions. From a business perspective, it also represents a loss of revenue and time for health information managers.

CHALLENGE

- Automate the most manual aspects of medical procedure coding.
- Support health information managers with billing code inquiries, insurance regulations, reimbursement guidelines, and other billing-related questions.

SOLUTION

Use today's software tools that use AI and machine learning algorithms to extract relevant information from patient data, such as electronic health records (EHRs).

- **Leverage these tools to generate accurate medical codes[9] and thus improve interactivity.**
- **Once coding is generated and validated, automate the process of sending bills to social security and insurance companies.**

RESULTS

- Avoid medical coding mistakes.
- Avoid income loss for hospitals and clinics.

ARTEFACT

USE CASE #5

Preventive and informational agent for patients

TARGET

Patients: As a patient, I receive medical care, treatment, or attention from a healthcare provider, such as a doctor, nurse, or hospital.

CONTEXT

In 2022, 52% of people[9] in the EU sought information on the Internet about their health or a medical procedure. And 70%[10] of them were exposed to misinformation without knowing it.

CHALLENGE

- Enable patients to retrieve reliable information ahead of a surgery or a medical procedure.
- Educate patients on medical conditions, medications and preventive care.

SOLUTION

- **Train an LLM with curated data sets describing medical procedures, preparation advice, medication posologies, adverse effects of medication, possible complications after surgery, and the main preventive reflexes to keep in mind by pathology.**
- **Set up a chatbot to provide patients with contextual and accurate responses to public health information[11].**
- **Enhance patients' knowledge about their overall hospitalization journey[12].**
- **Encourage patients to see a medical practitioner if the conversation reveals signs that care is needed.**

> ***We have not reached a stage where deploying Generative AI leads to job losses. While the technology holds immense potential in various fields, it often falls short in delivering results with the required precision.***
>
> **Vaibhav KULKARNI**
> Data Engineering Lead
> **DEBIOPHARM**

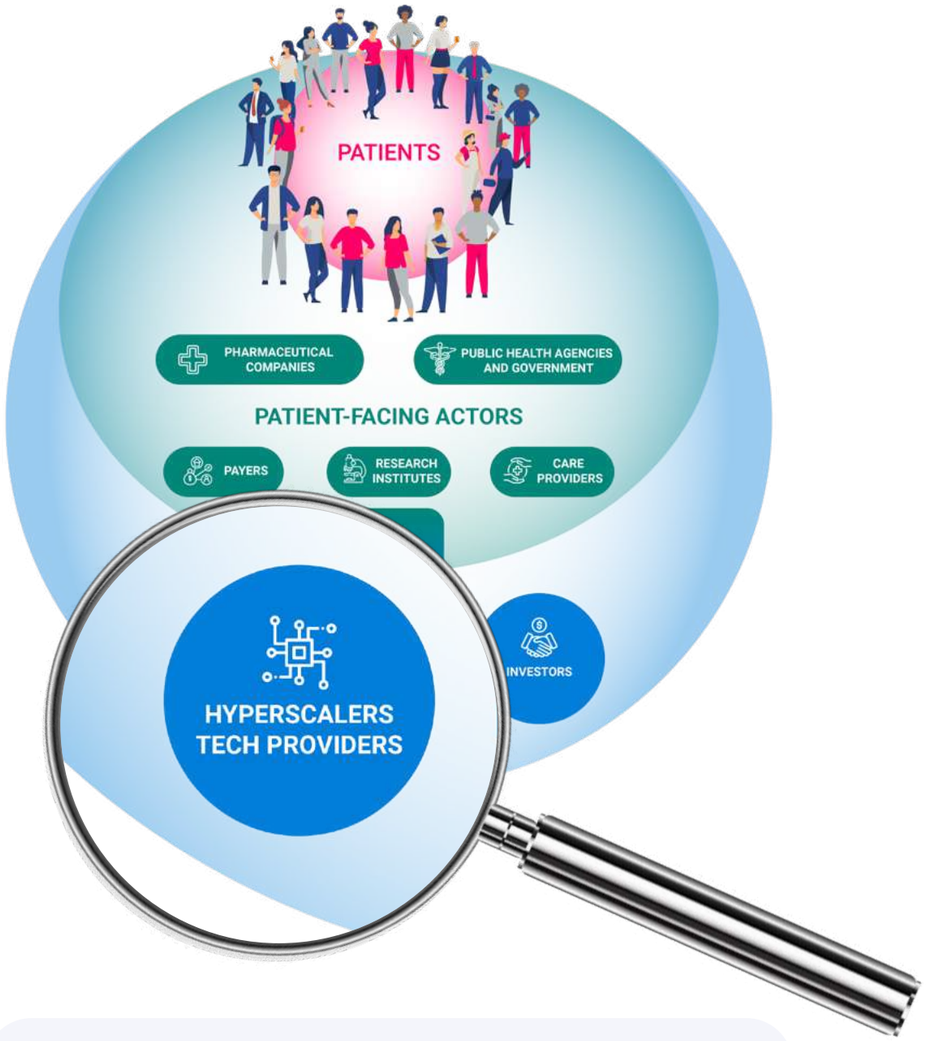
RESULTS

- Reduce stress linked to the hospitalization journey or any other medical treatment.
- Avoid misinformation by providing safe and private content, validated by health authorities.

3 The healthcare ecosystem: Gearing up to unlock the potential of Generative AI

Generative AI is a rapidly evolving technology and the healthcare ecosystem has already begun its metamorphosis in an attempt to harness its full potential. To better understand the progress of this transformation, we can divide the ecosystem into four main groups of players: hyperscalers, startups, pharmaceutical companies, and public domain actors.

1 — Hyperscalers and tech providers



amazon Google Meta Microsoft NVIDIA.

1.1 — Hyperscalers democratize Generative AI by building accessible models and tools

Hyperscalers successfully launched generalist foundation models available to the public and companies, starting with ChatGPT. Since its public launch in November 2022, it has captured the world's attention, showing millions of users around the globe the extraordinary potential of artificial intelligence. After their initial introduction to Generative AI, hyperscalers began taking on the challenge of fine-tuning these generalist models into domain-specific Large Language Models (LLMs). As healthcare is one of the areas with the greatest potential for transformation thanks to GenAI, it quickly became a top priority for hyperscalers.

Like everyone else, the healthcare industry first discovered Generative AI through publicly available models, such as OpenAI's ChatGPT or Google's Bard. The ecosystem reacted with enthusiasm, as HCPs and pharmaceutical companies' employees saw an opportunity to gain efficiencies in their everyday administrative operations.

To enable these administrative use cases, Big Tech companies developed applications such as Microsoft's Copilot, which consists of embedding OpenAI's ChatGPT into Microsoft 365. However, the healthcare ecosystem also quickly raised concerns around the unbridled use of ChatGPT or Bard in a pharmaceutical or healthcare environment: healthcare data is sensitive, and should not be shared in the publicly available versions of LLMs (*see part 3*).

To go beyond the basic augmented productivity use cases of LLMs and provide some reassurance with regard to data privacy, Big Tech created API services (GPT by OpenAI, PaLM by Google) to allow companies to configure their own generalist LLM while maintaining control of their data. Most LLM use cases we see emerging in the healthcare industry are derived from these APIs. Companies either fine-tune or prompt engineer these generalist models to fit their specific needs.

1.2 — Hyperscalers have already begun to create healthcare domain-specific Generative AI models and services to better address the complexities of medical data

Beyond the most administrative healthcare industry use cases, hyperscalers have already started creating healthcare domain-specific Generative AI models and services to better meet medical data complexity requirements. There is a lot of excitement around these initiatives, and while first results are promising, HCPs and pharmaceutical companies should be cautious about using these models which are far from limitless (see part 3). **We can look at four major examples:**

01 Med-Palm2 by Google

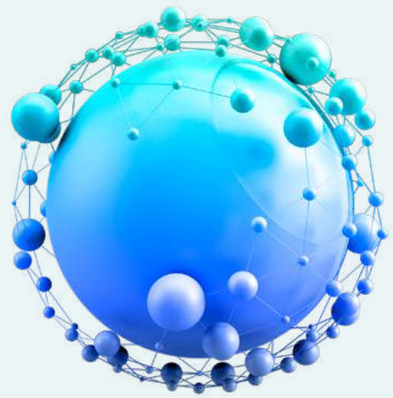
PaLM 2

Med-PaLM is an LLM designed to provide high-quality answers to medical questions. It was created by retraining Google's generalist LLM models with data such as medical exams, medical research papers, and healthcare consumer queries. Its first version was published in Nature in July 2023, and was the first AI to surpass the pass mark on US Medical License Exam (USMLE)-style questions. To date, Med-PaLM 2 achieves 86.5% accuracy on PubMedQA[13], the dataset used by the healthcare community to benchmark LLM performance.

Example of USMLE-style question

A 32-year-old woman comes to the physician because of fatigue, breast tenderness, increased urinary frequency, and intermittent nausea for 2 weeks. Her last menstrual period was 7 weeks ago. She has a history of a seizure disorder treated with carbamazepine. Physical examination shows no abnormalities. A urine pregnancy test is positive. The child is at greatest risk of developing which of the following complications?

A. Renal dysplasia
B. Meningocele
C. Sensorineural hearing loss
D. Vaginal clear cell carcinoma



02 Microsoft BioGPT

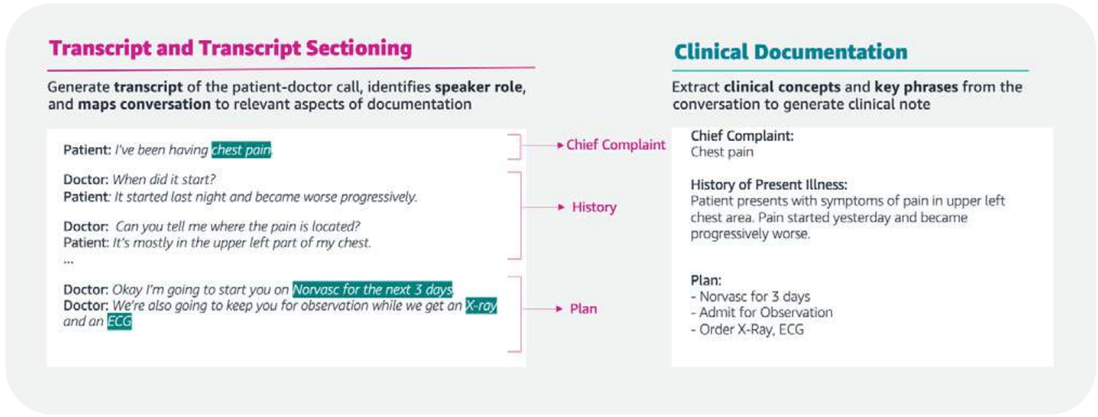
Microsoft

Microsoft has created an LLM called BioGPT based on GPT architecture that is specifically intended for processing text and biomedical data. It was trained on biomedical research articles so it can perform tasks such as answering questions, extracting relevant data, and generating text relevant to biomedical literature. For example, BioGPT can generate descriptions of a specific therapeutic class such as "Janus kinase 3 (JAK-3) or of a specific therapy—such as "Apricitabine." BioGPT achieved 81% accuracy on PubMedQA.

INPUT	BIOGPT OUTPUT
"Janus kinase 3 JAK-3)"	*"Janus kinase 3 (JAK-3) is a member of the Janus kinase (JAK) family of non-receptor tyrosine kinases and plays an important role in the regulation of cell proliferation, differentiation, survival, migration, and angiogenesis."*
"Apricitabine"	*"Apricitabine is an oral prodrug of 5-aza-2'-deoxycytidine (5-aza-CdR), a DNA methyltransferase (DNMT) inhibitor, which has been approved by the US Food and Drug Administration (FDA) for the treatment of myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML) in combination with low-dose cytarabine (Ara-C) and granulocyte colony-stimulating factor (G-CSF) for patients with intermediate-2 or high-risk MDS or AML"*

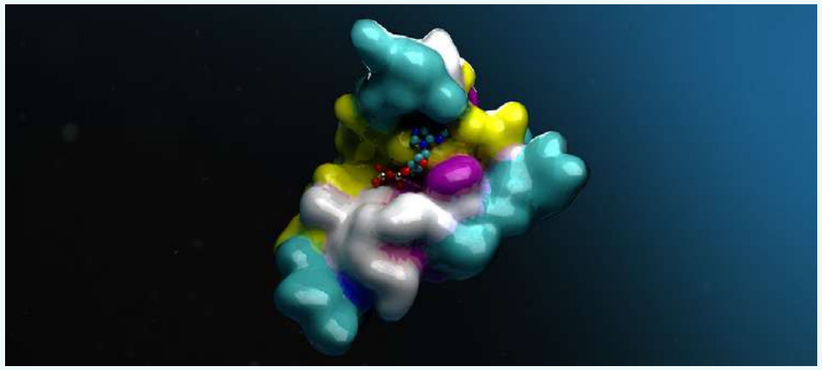
03 HealthScribe by AWS

HealthScribe is a clinical documentation service that enables healthcare software vendors to build clinical applications that use speech recognition and Generative AI to create transcripts of patient visits that auto-populate with additional relevant information, identify key details, and create summaries that can be entered into an electronic health record. It is currently being previewed for two specialties: general medicine and orthopedics.

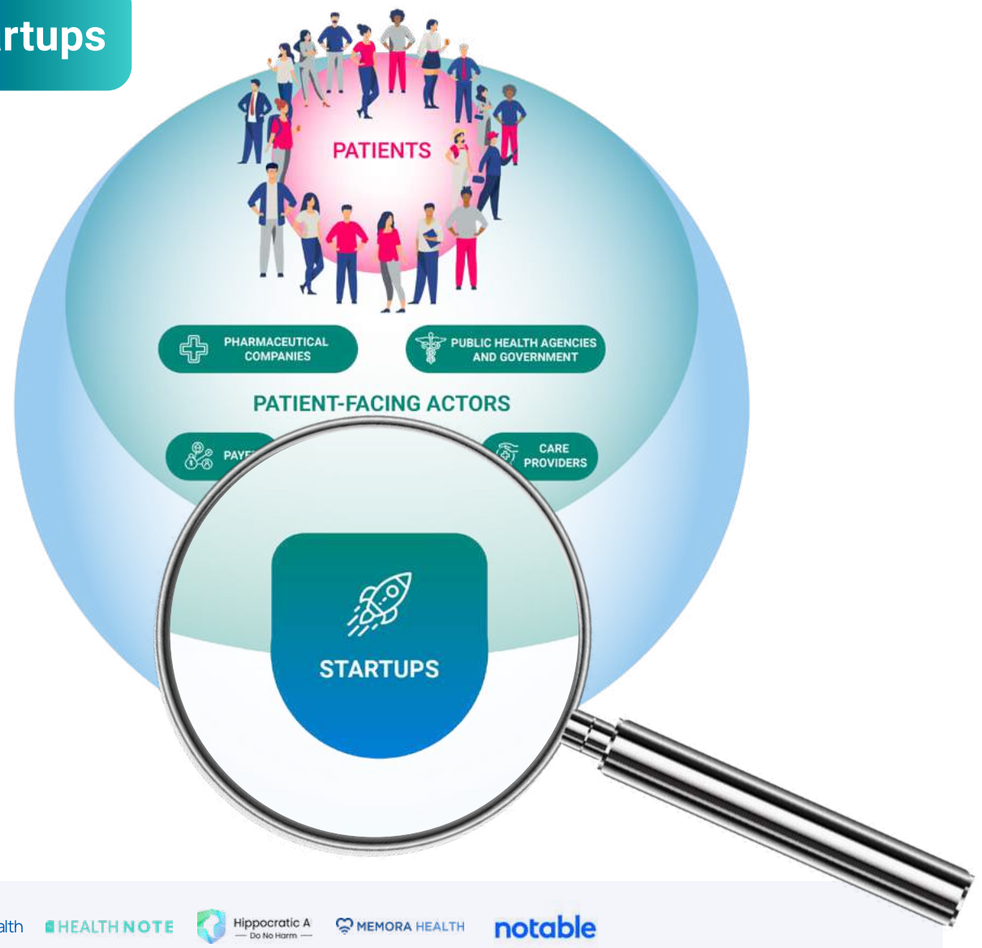


04 NVIDIA BioNeMo

In March 2023, Nvidia launched a Generative AI platform for drug discovery that enables researchers to fine-tune Generative AI applications on their own proprietary data, and to run AI model inference directly in a web browser or through new cloud application programming interfaces (APIs). BioNeMo offers 9 open-source foundation models with applications such as novel small molecule and protein sequence generation.



2 — Startups



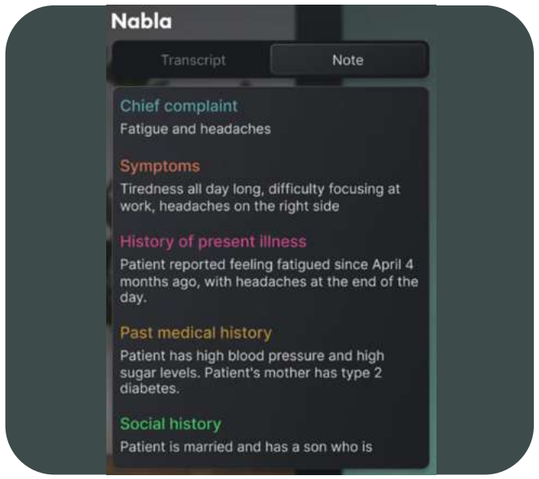
Assort Health · HEALTH NOTE · Hippocratic A — Do No Harm — · MEMORA HEALTH · notable
abridge · Ambience · Athelas · DeepScribe · Knowtex · Nabla · NUANCE
ATROPOSHEALTH · GLASS · OSTRO · REAL CHEMISTRY
Co:Helm · Develop Health · Develop Health · LATENT · tandem
CODAMETRIX · decoda health · FATHOM · nym · SmarterDx · regard
adonis · AKASA · candidhealth · Gentem · Health Harbor · JUNIPER NETWORKS
AQEMIA · Benevolent AI · IKTOS · Insilico Medicine

2.1 — Healthcare startups play a complementary role to hyperscalers by providing the ecosystem with off-the-shelf innovative solutions to address their more specific problems

While Big Tech excels at delivering large-scale solutions that cut across all industries, startups have demonstrated the ability to develop solutions that address the specific needs of the healthcare industry. With support from the pharmaceutical industry, venture capital and private equity investors, multiple disruptive startups have emerged to develop Generative AI applications around use cases that can deliver short-term value, such as HCP productivity, patient engagement and care decision making tools. To illustrate, we have selected three startups that offer off-the-shelf Generative AI modules:

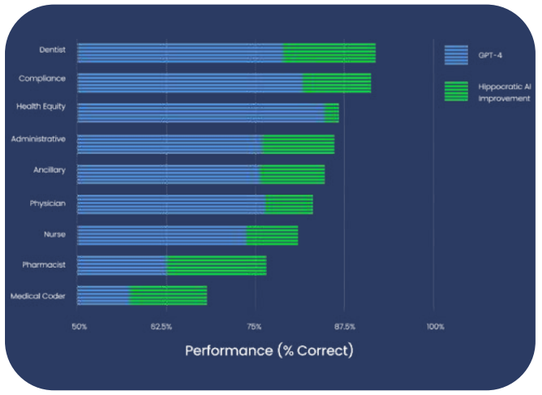
Nabla

Nabla Copilot combines speech-to-text and Generative AI technologies to offer a digital assistant for HCPs that can automatically transcribe consultations in real time and generate summarized clinical notes. It works for both in-person and video consultations. One of the key differentiating factors of the solution is its rapid deployment, which is achieved through three versions, all cloud-based: web-based app, Chrome extension, and API. In all cases, Copilot does not store or retain any data entered by the user. The API version can be integrated into the care center's electronic health record software. Based on publicly available information, Nabla is using OpenAI's GPT-3 LLM as the basis for Copilot, with the longer-term goal of building its own LLM, fine-tuned to the specific language and needs of medicine and healthcare.



Memora Health

Memora Health offers AI-powered care enablement that helps clinicians focus on delivering high-quality care while proactively engaging patients along complex care journeys. The platform relies on an LLM-enabled conversational AI capable of responding to routine patient messages. It operates on a retrieval-based approach, drawing on a selective but extensive clinician-validated database. This retrieval-based approach ensures that the AI only responds to patient questions with clinically-reviewed responses and only operates within clinician-validated pathways. More complex clinical concerns are automatically triaged to care teams.



Hippocratic AI

Hippocratic AI is building a safety-focused Large Language Model (LLM) designed specifically for healthcare, with an initial focus on non-diagnostic, patient-facing applications. This model is pre-trained on trusted, evidence-based healthcare content and already outperforms GPT-4 on 105 out of 114 healthcare exams and certifications. To ensure the safety of the model before it can be released to the public, the Hippocratic AI team is conducting a reinforcement learning with human feedback (RLHF) process using healthcare professionals to train and validate the model's readiness for use.

2.2 — Investors are joining the Generative AI revolution and have already begun financing high-impact projects

Private equity and venture capital firms are anticipating the impact of Generative AI in healthcare and have already started to invest in the technology. First, startups specializing in Generative AI are supported by an investment momentum strong enough to expand beyond the borders of their home country. Let's take another look at the French startup Nabla, which we mentioned earlier. Supported by Firstminute Capital and Artemis, they recently struck a deal with Permanente Medical Group in the United States, which is part of Kaiser Permanente, a healthcare giant with 75,000 employees. Nabla Copilot will initially be rolled out in Northern California, but if the product proves effective, it will be implemented throughout Kaiser Permanente's presence in the United States.

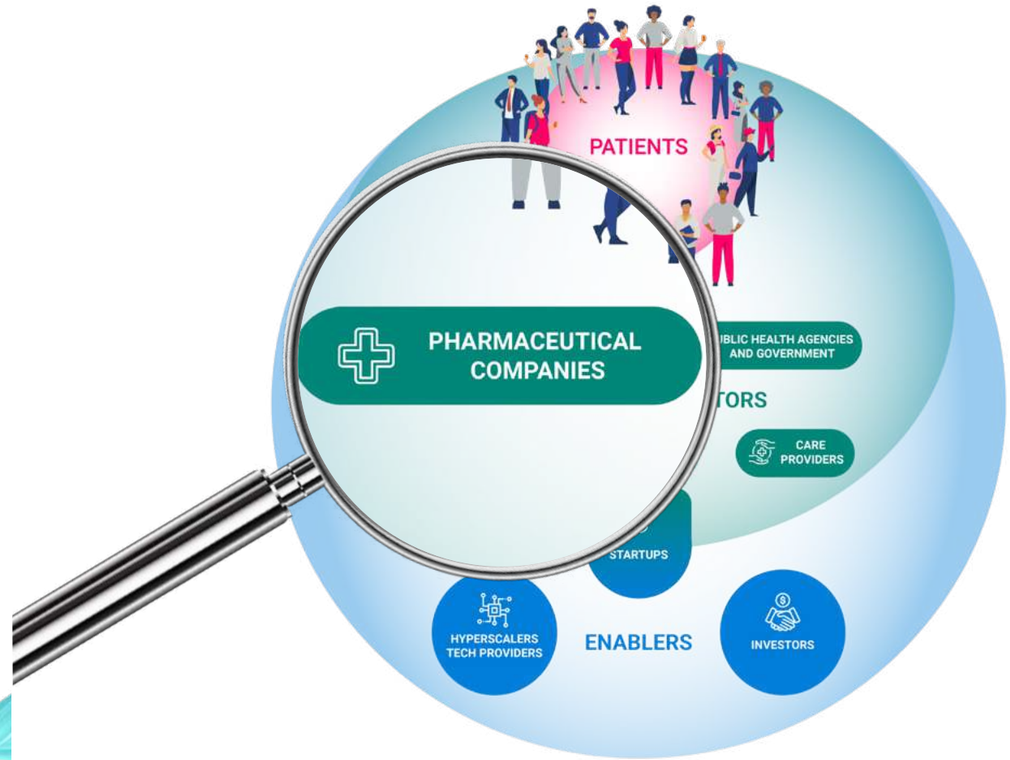
> *The challenge lies in integrating Generative AI into established companies that already have access to high-quality healthcare data, rather than investing in new startups.*
>
> **Anne-Sophie Saint-Martin -** Partner **- NEWFUND - Seed VC**

In addition to funding new Generative AI startups, investors are also positioning themselves to help established healthcare startups integrate this technology into their existing services. Ongoing projects focused on healthcare data processing, interoperability, or infrastructure are likely to explore Generative AI to augment their offerings. For example, Newfund, a French early-stage venture capital firm, is backing Arkhn and Omnidoc, two healthtech startups in its portfolio, in their efforts to integrate Generative AI into their current services.

Generative AI combined with quantum computing is opening new doors in drug discovery, and the transformative potential of this combination is attracting significant interest from investors. This is because the probabilistic nature of quantum computing can enable Generative AI to explore a broader solution space, potentially discovering novel treatments, optimizing healthcare processes, and uncovering patterns previously beyond the reach of classical computing. Elaia and other leading investment firms recently backed Aqemia, a next-generation pharmatech company using quantum computing to generate one of the world's fastest-growing drug discovery pipelines. The disruptive speed and accuracy of their technology platform allows them to treat and scale drug discovery projects as technology projects.

> *Generative AI is poised to generate numerous new opportunities in drug discovery. While most current projects are in the early stages of development, the combination of GenAI and quantum computing in drug discovery could not only lead to the creation of new treatments, but also to new advances that nature itself is not yet able to offer.*
>
> **Florian Denis -** Investment Director **- ELAIA**

3 — Pharmaceutical companies



3.1 — Large pharmaceutical companies have already started to develop proofs of concept (POCs) using the large language models available on the market

> *To ensure that the performance of AI models is maximized for all the patients we serve, we need to responsibly train these models on as many representative datasets as possible. When healthcare data is available at scale – in a way that protects patient privacy, first and foremost – we can create reliable databases and leverage the power of AI and machine learning to uncover insights that can help us better understand and measure disease onset and progression, design and optimize more targeted medicines, conduct more efficient and diverse clinical trials, and more.*
>
> **Tommaso Mansi -** Vice President of AI/ML & Digital Health
> **THE JANSSEN PHARMACEUTICAL COMPANIES OF JOHNSON & JOHNSON**

Pharmaceutical companies have a long history of using AI. They see Generative AI as a way to broaden its applications and, most importantly, to speed up the implementation of use cases and improve their ease of maintenance. To embark on an impactful Generative AI transformation and overcome the technical complexity inherent to medical data, most pharmaceutical companies would rather adopt a test-and-learn approach than rush into substantial investments towards industrialization. In fact, starting with an initial proof of concept enables them to quickly showcase concrete results and get people on board. This step is especially important as Generative AI is evolving on a daily basis, and pharmaceutical companies need to gain traction through results for the entire field, not just a specific application or technology. The healthcare ecosystem, as well as both French and European regulators, are still learning about the potential and limitations of Generative AI. And the specificity of the healthcare domain is that most Generative AI use cases, especially those that are HCP and patient-facing, have a very low tolerance for error.

Artefact already accompanies several pharmaceutical industry clients in the delivery of their first Generative AI POCs, for Research and Development, medical affairs, and marketing business departments.

3.2 — To keep pace with Generative AI advances in drug discovery, most pharmaceutical companies chose to enter strategic partnerships with key medical AI startups

Before the large language model hype, big pharmaceutical companies had already been experimenting with Generative AI in drug discovery and design for several years. To do so, most of them entered into strategic partnerships with key medical AI startups. Pharmaceutical companies invest in these innovative partnerships because traditional drug discovery and design has been an expensive and inefficient journey for too long. While it can take an average of 10 to 15 years to develop a drug, a 2023 article of the Drug Discovery Today magazine shows that a large pharmaceutical company spent an average of $4.4 billion annually on R&D and launched 0.78 new drugs between 2001 and 2020. Generative AI has already been shown to increase speed and efficiency at every stage of the process. It is being used to identify novel targets for disease and to design de novo molecules capable of acting on those targets. Some players are leveraging it to determine the likelihood of success in clinical trials.

3.3 — A few examples of innovative collaborations to watch

sanofi × Insilico Medicine

Insilico Medicine is a world-leading end-to-end Generative AI-driven biotech company with pipelines to explore transcriptome diversity by RNA-seq. They signed a strategic research collaboration with Sanofi in November 2022, worth up to $1.2 billion. Sanofi is leveraging their proprietary Pharma.AI platform across biology, chemistry, and clinical development to accelerate the discovery of novel therapeutics.

Pfizer × IKTOS

Iktos is an innovative company specializing in the development of artificial intelligence (AI) solutions applied to chemical research, specifically medicinal chemistry and new drug design. In March 2022, they announced the application of their Generative AI-driven de novo design software called Makya™ for de novo design to selected Pfizer small-molecule discovery programs.

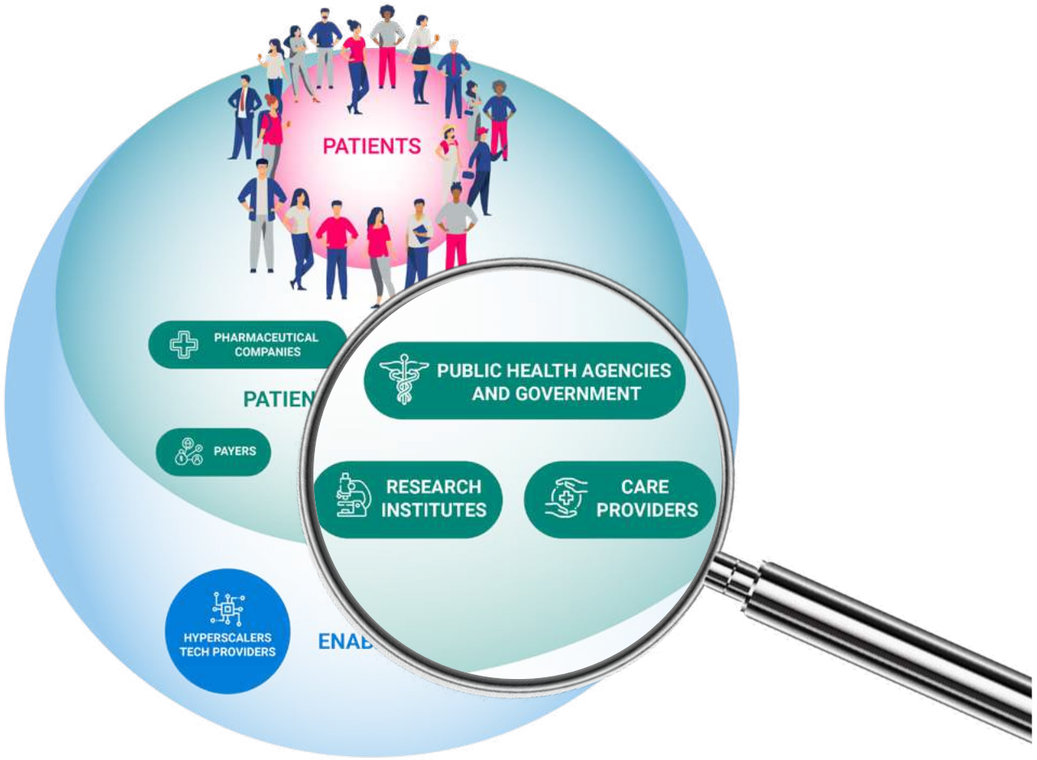
SERVIER × AQEMIA

Aquemia is a next-gen pharmatech aimed at designing fast innovative drug candidates for critical diseases. They rely on unique quantum algorithms and Generative AI to design novel drug candidates. Servier entered a partnership with them in December 2021 to accelerate drug candidate discovery in immuno-oncology using artificial intelligence and recently announced the extension of their collaboration on a new undruggable target in the above mentioned therapeutic area.

AstraZeneca × Benevolent AI

BenevolentAI is a clinical-stage AI and GenAI-enabled drug discovery company. Their key asset is the Benevolent Platform, a versatile, scalable and robust AI-enabled drug discovery platform built by expert scientists using multimodal data foundations. They also offer Generative AI products for knowledge exploration. AstraZeneca has been their partner since 2019, and this collaboration is bearing fruit as, based on BenevolentAI technology, the pharmaceutical company is advancing four of the most promising targets selected for its portfolio in chronic kidney disease (CKD) and idiopathic pulmonary fibrosis (IPF), while good progress is being made on further target selections in heart failure and SLE.

4 — Public domain



4.1 — Hospitals and research institutes have already begun to adapt this technology to their specific contexts

While some practitioners remain skeptical about Generative AI, the revolution it is bringing has not passed healthcare by. The healthcare community quickly realized that the generalist models made available by Big Tech companies would not be suitable for its needs. The use cases for Generative AI in healthcare have a much lower tolerance for errors compared to other industries. They also involve the processing of complex healthcare-specific queries, on which these models cannot perform well because they have not been trained on curated healthcare data. In France, healthcare stakeholders have therefore begun to develop their own foundational models, with the support of public funding entities.

To illustrate, we can look at the LLM4All Project, funded by the ANR (French National Research Agency) and led by public hospitals of Paris (AP-HP), CNRS, Computer Science Laboratory of École Polytechnique (LIX), and LINAGORA, an open source software and services company. In September 2023, the consortium announced its project aimed at providing retrained LLMs in open source to address some of the specific needs of the hospital sector. These models will be validated through two specific use cases in French: automatic summaries of SAMU meetings, and analysis of SAMU emergency calls. The call for projects, "Communs numériques pour l'intelligence artificielle générative", launched by the BPI in June 2023, will also be a major accelerator for Generative AI in healthcare in France. Several consortia of hospitals and medical research institutes have applied to develop new sovereign, open source foundation models dedicated to healthcare.

This call for projects launched by the BPI will also promote the creation of curated medical datasets for the purpose of training medical LLMs. The call for projects closed on October 24th, and while we do not yet have access to the full list of applications, we have learned through our panel of experts that key players such as the Health Data Hub, APHP, the University Hospital of Rennes (CHU Rennes) and the University Hospital of Reims (CHU Reims) have submitted large-scale applications in partnership with other hospitals and health research institutions.

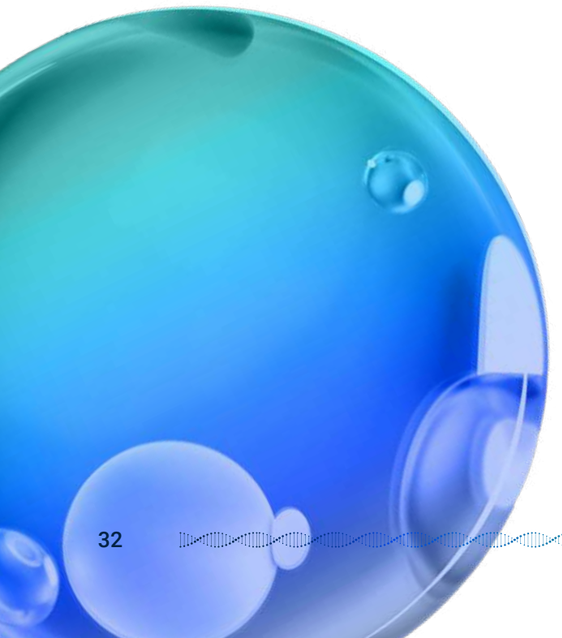
> *The AI revolution in healthcare is first and foremost a data revolution, not so much a modeling revolution.*
>
> **Emmanuel Bacry -** Senior Researcher CNRS, Chair PR[AI]RIE, Chief Scientific Officer **HEALTH DATA HUB**

To ensure the sovereignty of France and Europe in the field of Generative AI, public actors have started to equip themselves with supercomputers. Computing power is a limiting factor in the training of LLMs, and as these models are becoming larger and larger, supercomputers will be required to train them. Supercomputers are critical to a country's sovereignty in artificial intelligence, and even more so in Generative AI. They enable advanced AI research and model training, which is essential to remain competitive globally. One of France's main research organizations, the CEA, recently announced that it will host the future European Exascale supercomputer as of 2025, which will be one of the fastest supercomputers in the world. This supercomputer will bring Europe back into the supercomputing race and will act as a sovereign accelerator for several strategic European data-related challenges, such as training the next generation of Generative AI or multimodal models.

4.2 — Foundation LLMs provided by hyperscalers have one major limitation for the French healthcare ecosystem: they are not sovereign. In addition to emerging initiatives by French hospitals and medical research teams, Docaposte, a French expert in sensible computing, recently announced the launch of its first sovereign LLM service with healthcare use cases.

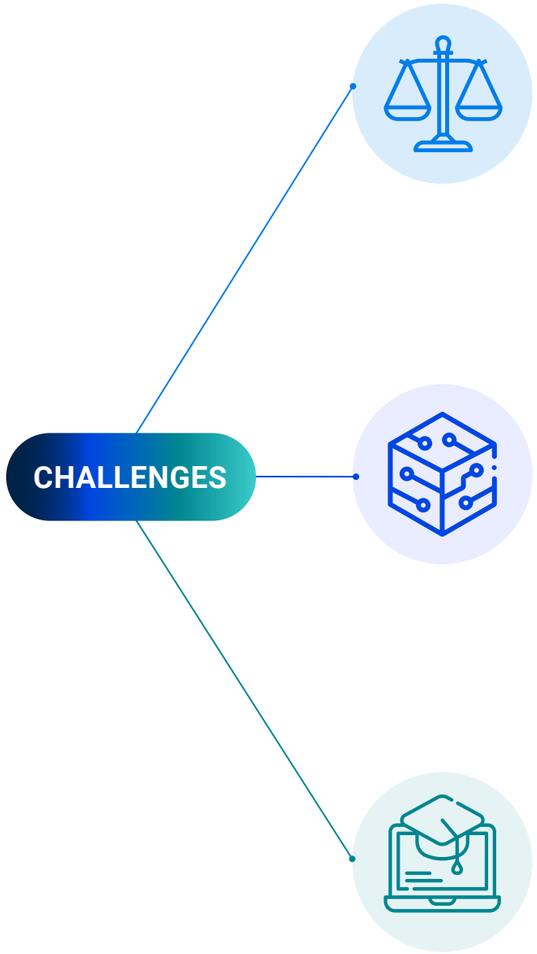
DOCAPOSTE

In partnership with three French companies (LightOn, Aleia, and NumSpot), Docaposte is launching its first sovereign, cloud-based LLM, available from November 2023. This offering is aimed at French organizations that handle sensitive data, particularly in the healthcare sector. Docaposte unveiled a first application called "MyAssistant Santé". They trained the foundation LLM called Alfred, provided by LightOn, using the reference documents that govern the patient's medical record. The resulting model powers a chatbot that assists the doctor in gathering information about a patient's case before treatment and more effectively transcribes information back into the same patient record.



4 Generative AI in healthcare: Limitations, challenges and opportunities

Although Generative AI holds the promise of revolutionizing the healthcare industry, it comes with its own set of risks and challenges that need to be addressed.



Legal risks

Considering the ongoing uncertainty surrounding the full extent of this technology's implications, it is crucial for industry stakeholders and users to engage in meticulous deliberation, especially when addressing the legal aspects in the healthcare sector. These legal aspects include issues such as liability for medical errors, limitations on data flow due to privacy and confidentiality regulations, and the crucial question of data sovereignty for data owners.

Data and technology

From a technological point of view, Gen AI faces several key challenges. On one hand, it relies heavily on deep learning techniques to produce new content, leading to a certain opacity in its decision-making process and a risk of biases emerging. On the other hand, its effectiveness depends on the availability of a large amount of data which, in the medical field, is often difficult to acquire due to privacy concerns and data access restrictions, therefore limiting its capabilities at present.

Training and adoption

It is imperative for institutions and companies to implement well-structured change management processes in order to integrate AI into their operations. Equally crucial is the need for comprehensive education and awareness programs for healthcare professionals and C-suite executives to provide them with a deep understanding of both the capabilities and limitations of AI, as well as an awareness of its vulnerabilities. This knowledge is essential to effectively mitigate and balance the inherent risks associated with the use of AI in healthcare.

However, these challenges also open up new opportunities for disruptive players to leverage their existing assets and expand their offerings into the realm of Generative AI.

This section, based on the interviews we conducted and our experience in delivering client projects, highlights the most significant risks related to Generative AI in healthcare. It also explores potential strategies for mitigating these risks to ensure responsible and ethical implementation, along with the opportunities that arise.

1 — The need for robust legal and regulatory frameworks to safeguard health data privacy, security, and integrity

Given the highly sensitive nature of healthcare data in terms of confidentiality, it is inherently subject to a variety of privacy regulations and requirements. As artificial intelligence and machine learning become more integrated into the field of medicine, there is a growing concern that inaccuracies in algorithms could lead to patient injury and medical liability.

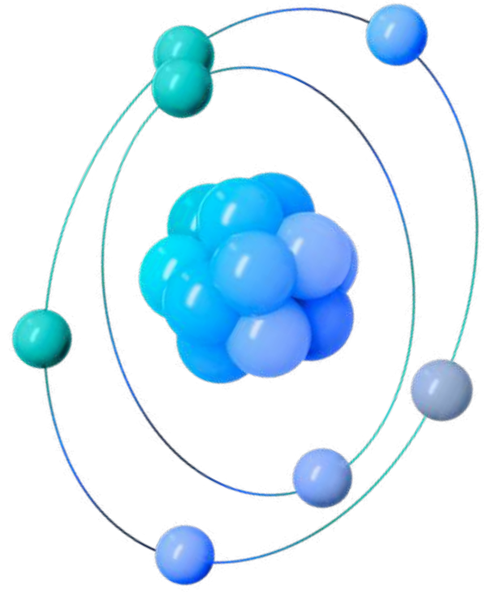
While certain regulations have already been put in place to protect patients' personal data, such as the General Data Protection Regulation (GDPR) in Europe and the Health Insurance Portability and Accountability Act (HIPAA) in the United States, both national and international regulatory bodies are actively working to keep pace with the rapid advances in Generative AI.

One notable example is the European Union's proposed AI Act, the world's first concrete initiative to ensure that AI systems deployed in the EU are safe, transparent, traceable, non-discriminatory, and environmentally friendly. It proposes a principle of graduated rules by classifying AI models according to the risk they pose to users (minimal, limited, high or unacceptable risk), with associated obligations and counterbalancing measures (from minimal transparency requirements to impact assessment studies and measures to mitigate potential risks).

While the primary objective is to safeguard the interests of citizens against unethical use of their data, a commendable goal, it inadvertently poses challenges to data access and sharing. The regulations aim to strike a balance between privacy and technological advancement. Will Europe lag behind other regions in the development and adoption of Generative AI, or will it set a precedent for future global rules and standards in this rapidly evolving field? It is too early to predict what this regulation will mean for the future.

2 — Ensuring privacy and safety around healthcare data and Gen AI models underscores the need for innovative tools

While data anonymization is the most common approach for safeguarding privacy, it can be a complex task for medical data sources. The main challenge does not lie in the removal of overtly personal information such as names or social security numbers, which can be readily redacted or randomized. The true complexity emerges when addressing indirect data points, such as social connections (often referenced in medical reports), intricate medical histories, or nuanced spoken language. While these elements may not be distinctly identifying on their own, their collective amalgamation can yield powerful identifiers capable of describing a unique individual or an exceedingly small cohort.



3 — Sarus, an innovative start-up addressing privacy issues

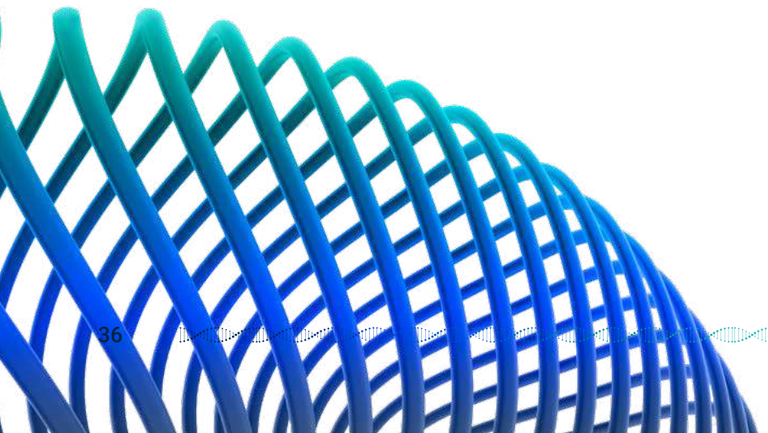
To circumvent the challenges of anonymization, Sarus offers LLM fine-tuning powered by differential privacy. It ensures that no personal information is embedded in the fine-tuned LLM. In short, Differential Privacy (DP) is a mathematical approach that protects personal information during data processing. This goal is achieved by adding controlled noise to the computations, making it virtually impossible to identify individual data points. This noise is adjusted to protect privacy while still allowing for valuable statistical analysis. This methodology protects patient data without the need for anonymization.

4 — At present, the ability to explain the decision-making process of an AI remains elusive, making the explainability of LLM models a significant challenge

Embracing AI is not just about integrating the technology, but also understanding its decisions while mitigating biases. Algorithms are often referred to as “black boxes” because of their ability to evolve as new data is integrated and because they rely on deep learning algorithms to create new content. In addition, there is growing evidence that these algorithms can sometimes produce recommendations that are racially or otherwise biased.

Bias awareness is crucial, as most users struggle to distinguish between truth and the likelihood of a result. While truth stands as an unambiguous and factual representation - essential for a definitive medical diagnosis or assessment - likelihood operates within the realm of probabilities. Indeed, AI generally relies on statistical inference and its output is determined by what is “most likely”, given its training data, algorithms, and the nature of queries. In particular, LLMs often generate responses that may appear authoritative and plausible to end users, yet they may be completely incorrect or contain serious errors. This phenomenon known as “hallucination” is a common bias inherent in LLMs. In a sector such as healthcare, it is imperative that this problem is addressed at its source, as it could pose a risk to patients. While some solutions (such as Retrieval Augmented Generation (RAG), fine-tuning LLMs…) already exist, their widespread implementation remains a work in progress.

In fact, the World Health Organization (WHO), which published the Ethics & Governance of Artificial Intelligence for Health in 2021, recently called for caution because “the data used to train AI may be biased, generating misleading or inaccurate information that could pose risks to health, equity, and inclusiveness.”



> ***When prompted about some topics they don’t know, LLMs tend to “hallucinate” wrong responses. One way to mitigate this phenomenon is to automatically retrieve the documents in a knowledge base that are most likely to contain elements of the response and add them to the prompt so that the LLM has more context to give a correct answer. This technique is called Retrieval Augmented Generation (RAG). Other techniques can be used depending on the use case, such as: fine-tuning. When dealing with confidential data, these approaches should be combined with techniques to prevent the leakage of confidential data in LLM responses.***
>
> **Nicolas Grislain -** Co-founder and Chief Scientific Officer **- SARUS**

5 — LLMon, an innovative approach for detecting AI Safety risks in LLM applications

The solution, developed by French startup Giskard, aims to provide a robust defense against potential pitfalls such as hallucinations, ethical biases, and inaccuracies that can occur in LLM-based applications. By verifying that outputs are accurate, it is designed to be a steadfast companion, offering comprehensive monitoring and evaluation capabilities for LLM-driven endeavors.

Even with the development of AI security-enhancing tools, it remains crucial for both professionals and patients to understand the difference between having confidence in an answer and granting unconditional trust. This nuanced understanding serves as a catalyst for more informed decision-making and acts as a safeguard against overreliance on AI, emphasizing the essential role of human oversight.

6 — In addition to technology, data accessibility plays a pivotal role in providing the necessary input for the operation of Generative AI models

The significant progress made by Generative AI depends primarily on the accessibility of the data needed to fuel these emerging models. Currently, a major challenge is the collection of large amounts of high-quality data to ensure reliable outcomes. Even in an era characterized by data ubiquity, obtaining top-quality data in the healthcare sector faces various technical and ideological obstacles. The scarcity of healthcare data can hinder the development and training of Generative AI models, which rely on large and diverse datasets to learn effectively and produce accurate and reliable outcomes.

Another phenomenon exacerbating the challenge of data accessibility is the distinct concern about data monetization that has emerged, particularly in the French landscape. While the trend is moving towards a greater willingness to share data, the difficulty of evaluating the financial value of the services associated with data and/or databases was further accentuated during the COVID-19 pandemic, when data sharing was needed more than ever. Paradoxically, this period witnessed the reinforcement of a "data ownership" fantasy, which ran counter to the expectation that it would diminish precisely when the demand for collective data sharing was at its peak.

> ***As access to healthcare data is particularly difficult, one of the key differentiators for organizations wishing to embark on a Generative AI journey will be their ability to access large quantities of data.***
>
> **Benjamin Belot -** Partner **- KURMA PARTNERS**, VC focusing on Healthtech & Biotech

7 — The ecosystem must find creative solutions to address data scarcity while respecting privacy and regulations

Key stakeholders in the healthcare ecosystem who have not yet fully embraced Generative AI bear a significant responsibility in providing high-quality medical data for training Large Language Models (LLMs), such as medical software editors like Cegedim, Doctolib or Osiris. Moreover, healthcare-centric AI companies like Owkin, are already collaborating with numerous hospitals and research institutions worldwide through their federated research network. Medical data warehouse providers like Arkhn or Codoc also have the potential to leverage their network of hospital partners to provide both structured and unstructured data for training LLMs.

In addition, another way to overcome the scarcity of data challenge is to rely on GenAI's ability to generate synthetic data, such as medical images. This approach allows training data to be augmented, thereby promoting diversity in medical research and training. It mitigates privacy concerns by removing sensitive patient information while retaining the essential features needed for meaningful analysis. It thus provides an answer to the common problem of data scarcity and confidentiality, facilitates data sharing, and improves model robustness.

8 — Data acculturation and training are indispensable to familiarize healthcare professionals with the use and potential associated risks of this technology

A lack of public awareness of biases in data, models, and their applications can lead to potential misinterpretation of the tool, its scope, and its results.

Institutions and companies should prioritize the implementation of comprehensive change management processes for the seamless integration of AI. This entails more than just installing software; it involves a fundamental reevaluation of the organization's core values. It includes educating collaborators about AI's capabilities and constraints, while highlighting its pivotal role in achieving the organization's objectives, such as enhancing patient care and providing time-saving tools.

To ensure the genuine effectiveness of AI, including GenAI, it is imperative not only for the technology to be state-of-the-art but also for the professionals using it to be highly skilled. This is especially critical for healthcare professionals who are primary users and directly interact with patients. Currently, there is a notable lack of comprehensive training programs that are tailored to different user proficiency levels, from novice to expert. Such programs are essential for making professionals not only operationally proficient, but also culturally sensitive to the nuanced aspects of AI.

> ***It is essential to frame the use of a Generative AI model with a usage convention and to ensure in-depth training for health professionals regarding its intrinsic constraints and vulnerabilities.***
>
> **Jean-Marc Bereder -** Artificial Intelligence Usage Specialist and Former Head of department **- NICE UNIVERSITY HOSPITAL**

> ***Over the long term, humans risk becoming overly reliant on generated documents, which could lead to a loss of comprehension and technical skills. Human decision-making must be preserved to prevent alienation caused by LLMs.***
>
> **Vincent Vuiblet -** Professor of Universities and Hospital Practitioner **- CHU REIMS**, URCA and Director of l'Institut d'Intelligence Artificielle en Santé Reims Champagne Ardenne (I2AS)

5 Trust and control play a critical role in realizing the potential of Generative AI in healthcare. It should therefore be seen as a human transformation, not just a technical one.

Generative AI is a game changer for the healthcare industry, transforming the way we interact with data, and consequently how we analyze, make decisions, and personalize our interactions. This year, we have observed an extremely rapid increase in the awareness among all players in the ecosystem: tech players are developing specific enablers to better address healthcare challenges, while research institutions, industry players, caregivers, and public institutions have mostly initiated consultations and initial projects.

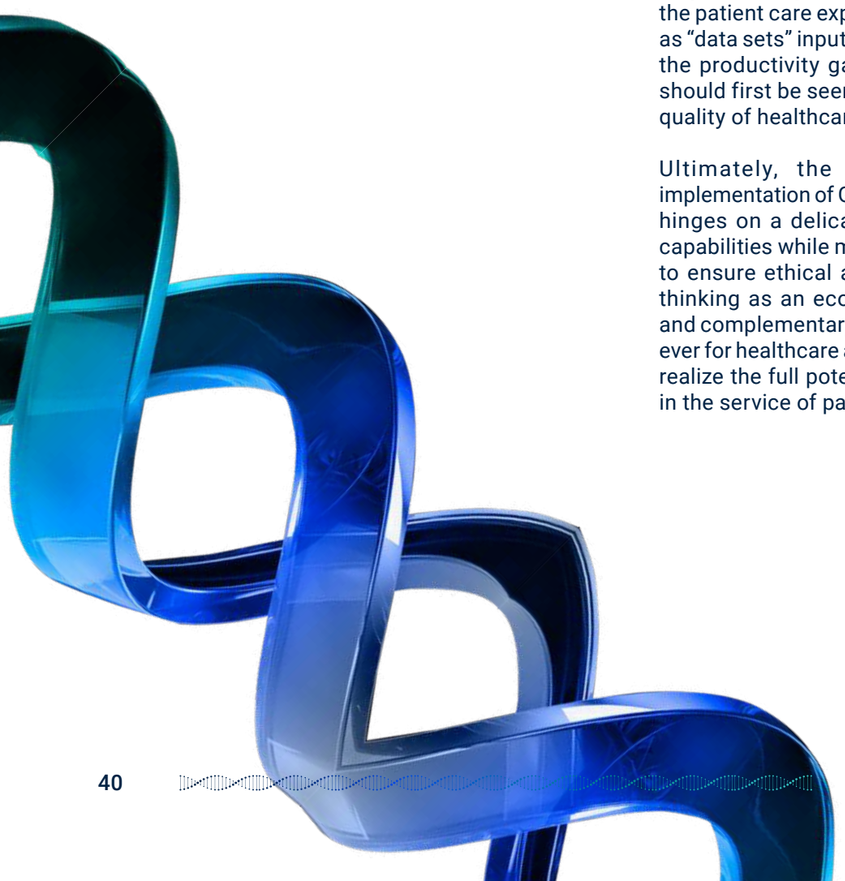
However, this enthusiasm should be tempered as the technology is still too new and immature. Many players are moving forward knowing that the potential is significant, but with a high degree of uncertainty about the achievable performance, the possible level of industrialization, and the regulatory constraints that will be defined.

To overcome these challenges, it is essential that healthcare and pharmaceutical industry stakeholders keep humans at the center of the Generative AI transformation at all times.

First, humans should be at the center of decision-making processes. Individuals should not only be trained in the use of these technologies but also equipped with the tools and capabilities to exercise control and make informed decisions. Indeed, the primary objective of AI in healthcare is to empower the medical ecosystem, ensuring that technology serves as an enabler of human expertise rather than a replacement for it.

Second, we must ensure that humans remain the primary beneficiaries of this transformation. Indeed, the widespread diffusion and adoption of these technologies, if perceived only as productivity tools, could lead to a very technological view of health, thus depersonalizing the patient care experience by perceiving patients only as “data sets” inputs and outputs for models. However, the productivity gains of Generative AI applications should first be seen as a means to improve the overall quality of healthcare and patient management.

Ultimately, the responsible and sustainable implementation of Generative AI in the healthcare sector hinges on a delicate balance: harnessing AI for its capabilities while maintaining strong human oversight to ensure ethical and responsible use. Building and thinking as an ecosystem with common challenges and complementary strengths, is more necessary than ever for healthcare and pharmaceutical stakeholders to realize the full potential of data, AI, and Generative AI in the service of patients.



Thanks & acknowledgements

We would like to express our sincere gratitude to all the interviewees from all parts of the ecosystem (researchers, hospital practitioners, startups, pharmaceutical companies, investors, attorney) who contributed to the creation of this white paper. Their expertise and collaborative spirit were instrumental in shaping and refining our ideas.

Pr Stéphanie Allassonnière - Professor and Vice-President Valorisation and industrial partnerships, Chair PR[AI]RIE - **Université Paris Cité**

Emmanuel Bacry - Senior researcher CNRS, Chair PR[AI]RIE - Chief Scientific Officer - **Health Data Hub**

Benjamin Belot - Partner at Kurma Partners - VC focusing - **Healthtech & Biotech**

Jean-Marc Bereder - Artificial Intelligence Usage Specialist and Former Head of department - **Nice University Hospital**

Marguerite Brac de la Perrière - Partner Attorney at Law in Digital Health - **Fieldfisher**

Thomas Clozel - CEO and Co-founder - **Owkin**

Florian Denis - Investment Director - **Elaia**

Nicolas Grislain - Co-founder and Chief Scientific Officer - **Sarus**

Vaibhav Kulkarni - Data Engineering Lead - **Debiopharm**

Tommaso Mansi - VP of AI/ML & Digital Health - **The Janssen Pharmaceutical Companies of Johnson & Johnson**

Dr Grégoire Pigné - CEO - **PulseLife** - Oncologist and Radiation Therapist

Anne-Sophie Saint-Martin - Partner - **Newfund** - Seed VC

Pr Eric Vibert - MD, PhD, IKO, Liver Surgeon in Paul Brousse Hospital AP-HP, Paris Saclay University - **Chairman of BOPA**

Pr Vincent Vuiblet - Professor of Universities and Hospital Practitioner - **CHU Reims and URCA** - Director of Institut d’Intelligence Artificielle en Santé Reims Champagne Ardenne (I2AS)

We also extend our heartfelt thanks to each member of our team for their tireless efforts and invaluable contributions throughout the entire process.

Artefact editorial team:

Paul de Balincourt - Director - Healthcare Practice - **ARTEFACT**

Tanguy Masgnaux - Senior Consultant - Healthcare Practice - **ARTEFACT**

Sébastien Marguerès - Programs & Scientific Manager - **AI for Health**

Léa Giroulet - Senior Consultant - Healthcare Practice - **ARTEFACT**

Emma Tordo - Consultant - Healthcare Practice - **ARTEFACT**

Artefact sponsors of the initiative:

Justine Nerce - Managing Partner - Healthcare Practice Lead - **ARTEFACT**

Damien Gromier - Founder & CEO - **AI for Health**

Stéphanie Trang - General Manager - **AI for Life**

01 — Clinical trials phases description: Phase I evaluates the tolerance of a human body after the administration of the drug. Phase II evaluates the efficiency of the drug. Phase III ensures at a larger scale the efficiency and safety of drugs.
02 — “The minimum threshold is usually set to 300” - Stéphanie Allassonnière, 2023
03 — The retention of patients in the whole clinical trial duration can change for various reasons (distance, understanding, health complications,...), leading to some data gaps or lack throughout the process.
04 — Patient journey data (not exhaustive): it includes past surgeries, comorbidities, medication and duration of treatment as well as the age, gender, height, and weight of the patient.
05 — Synthetic Data Generation by Artificial Intelligence to Accelerate Research and Precision Medicine in Hematology
06 — HAS website
07 — Purpose of these Institutions, through the standards, is to guarantee the quality and safety of care for patients
08 — Implement Gen AI in medical acts coding
09 — Eurostat
10 — AHIMA Foundation: Spotting health misinformation online
11 — Example of the Covid-19 chatbot.
12 — Example addressed by Pr. Vincent Vuiblet: patients will be able to increase their knowledge about their hospitalization, to understand each step (before, during, and after the hospitalization).
13 — MultiMedQA is a comprehensive collection of multiple-choice medical question-answering datasets, used for training and evaluating Med-PaLM. MultiMedQA is comprised of the following datasets: MedQA, MedMCQA, and PubMedQA. PubMedQA is usually the data set used for evaluation.

Glossary

Artificial Intelligence (AI): The field of computer science that focuses on creating intelligent machines capable of performing tasks that typically require human intelligence, such as visual perception, speech recognition, problem solving, and decision-making.

Machine learning: A subset of AI that involves developing algorithms and models capable of learning patterns and making predictions or taking actions based on data, without being explicitly programmed.

Deep learning: A subset of machine learning that utilizes artificial neural networks with multiple layers to process and learn from large amounts of data, enabling the model to make complex and sophisticated predictions or decisions.

Foundation model: A pre-trained and highly capable AI model that serves as a basis or starting point for developing more specialized models, enabling faster development and reduced training time for specific tasks or domains.

Generative AI: A branch of artificial intelligence (AI) that focuses on creating and generating new content, such as images, music, or text, using algorithms and models.

Large Language Model (LLM): A powerful AI model capable of processing and generating human-like text, leveraging a vast amount of pre-existing language data to generate coherent and contextually relevant responses.

AI Act: Proposed regulation by the European Union that aims to be the world's first concrete initiative for regulating AI by establishing a legal framework for the development, deployment, and use of AI systems.

AI bias: A phenomenon that occurs when an AI algorithm produces results that are systemically prejudiced due to erroneous assumptions in the machine learning process.

Clinical trial: A type of research that studies new tests and treatments and evaluates their effects on human health outcomes.

Cohort: A group of people who share a particular characteristic, such as social and health factors.

Differential privacy: A mathematical technique that adds a controlled amount of randomness to a dataset to prevent anyone from obtaining information about individuals in the dataset.

General Data Protection Regulation (GDPR): A comprehensive data protection regulation in the European Union (EU) that governs the collection, processing, and storage of personal data, and aims to protect the privacy and rights of individuals.

Healthcare Professional (HCP): Any natural person who is a member of the medical, dental, pharmacy or nursing professions, or any other person who, in the course of his or her professional activities, may prescribe, purchase, supply, recommend or administer a medicinal product

Hyperscalers: Large cloud service providers that offer services such as massive computing resources and storage at scale (AWS, Microsoft Azure, Google Cloud Platform…)

Model fine-tuning: The process of adjusting and optimizing a pre-trained AI model by exposing it to additional data, specific to the desired task or domain, in order to enhance its performance and make it more suitable for the target application.

Programming language: A formal language with a set of rules and syntax used to write computer programs. It provides the necessary instructions to tell a computer what tasks to perform, and it can be used to develop applications, algorithms and AI models.

Prompt engineering: The process of designing and optimizing prompts or instructions given to a language model with the goal of eliciting specific and desired outputs or responses from the model.

Real-world evidence (RWE): Clinical evidence of the safety and efficacy of a medical device generated using real-world data (RWD) from routine health care delivery.

Retrieval Augmented Generation (RAG): An AI framework for improving the quality of LLM-generated responses by grounding the model on external sources of knowledge to supplement the LLM's internal representation of information.

Synthetic data: Information that's artificially created rather than generated by real-world events to augment or improve AI models.

Tech providers: Company or organization that offers various technology solutions, products, or services for specific aspect(s) of an industry.

ARTEFACT

DATA IS ABOUT PEOPLE

CONTACT

hello@artefact.com
+ 33 1 79 72 45 45
artefact.com/contact-us

ARTEFACT HEADQUARTERS

19, rue Richer
75009 – Paris
France

artefact.com

ABOUT ARTEFACT

Artefact is a global leader in data & AI consulting and data-driven marketing services, dedicated to transforming data into business impact and tangible results across the entire value chain of organizations.

Artefact's skyrocketing growth is fueled by our visionary and entrepreneurial founders as well as a unique methodology and multidisciplinary teams.

Artefact helps clients put consumers at the heart of their strategy through a comprehensive range of data-driven solutions. Designed to meet the specific needs of today's clients, our services are delivered via a business-centric approach that is built upon deep AI expertise. We are a connected independent global network and we partner with 1000+ ambitious clients around the globe.

Our 1300+ employees, present in 18 countries (Europe, the Americas, Asia, Middle East Africa) are focused on accelerating digital transformation thanks to a unique mix of company assets: cutting-edge AI technologies, agile methodologies for fast delivery and efficient scalability, and teams of market-leading experts in data science and digital marketing, always working together and focusing on business innovation.

IN PARTNERSHIP WITH

AI FOR HEALTH
BY ARTEFACT